DOCUMENT RESUME

ED 131 345 CG 006 215

AUTHOR Hazlett, James A.
TITLE National Assessment--A Tool for Accountability.
PUB DATE 23 Feb 71
NOTE 16p.; Paper presented at the Annual Meeting of the American Association of School Administrators (February, 1971)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Data Analysis; *Educational Accountability; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Social Science Research; Speeches; State of the Art Reviews; Statistical Studies
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

James A. Hazlett, NAEP Administrative Director, discusses the National Assessment program as an accountability tool. The strengths and weaknesses of the program as an evaluative method in the cognitive areas are discussed at length. (MPJ)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED131345

National Assessment--A Tool for Accountability

Remarks of James A. Hazlett - NAEP Administrative Director
AASA Section Meeting -10:00 a.m. Monday, February 23, 1971

The National Assessment Program was conceived as a tool for accountability several years before accountability came into prominence as an educational responsibility. While the concept of accountability means different things to different people, there are three aspects to the concept which seem consistently discussed: (1) the evaluation of pupil performance and behavior, (2) the identification of factors in the teaching-learning situation which have a direct bearing on performance and behavior, and (3) an attempt to relate cost of these factors to learning outcomes. As an accountability tool National Assessment does not involve itself in all three elements; but for that matter no accountability tool of present does do the complete job, although some theoretical models are designed to consider all these relationships.

It probably is fair to say that the great interest in accountability today is the product of two developments of the past decade: (1) allocations to education by state, local, and federal governments to a degree that has exceeded allocations to other public services at a time when rising taxes, especially on real property, are not only being questioned but are being rejected in many instances by taxpayers and legislatures, and (2) a governmental and public awareness that large segments of the school population, particularly minority groups, are educationally deficient in attaining acceptable standards and are dropping out of school in large numbers.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 006 215

Real questioning of the effectiveness of the educational enterprise has resulted in an array of issues and suggestions for improvement of educational quality and is seriously challenging the entire institution. Among these suggestions and demands are (1) change and innovation in teaching methods and curriculum, (2) more direct community control of schools, (3) performance contracting, (4) program budgeting, (5) the voucher system, and (6) better assessment of needs and results. The motivation behind all these movements is the conviction that schools can do better and must demonstrate that they can.

Any effective system of accountability must be based upon (1) a statement of explicit objectives which are understood and agreed upon, (2) the employment of evaluation schemes which can measure pupil performance in terms of these objectives, and (3) hopefully relate costs of instructional elements to outcomes.

As I see the various programs advocated to improve schools, PPBES, or program budgeting, is the one most comprehensive conceptually in considering all the elements required for accountability, followed next by some of the developments in state assessment planning. Some of the programs just mentioned are motivated by a desire to make schools more accountable but lack a methodology to determine if the program would actually, in fact, improve quality.

Our purpose here this afternoon is to analyze National Assessment as an accountability tool and to look at its strengths and limitations. My experience with the project over the last year and a half tells me that National Assessment is not well understood, that too many people are expecting too much from it, and that it has many elements designed to improve educational excellence that are not fully appreciated.

Let's first of all look at its strengths.

Strengths

1. First of all, National Assessment stands fully behind the proposition that schools and school systems can best be evaluated by looking at the results of educational effort rather than by judging the factors that make up the teaching-learning situation against arbitrarily set standards; that quality is measured by pupil performance and behavior, rather than by class size, the training of the teacher, or per pupil expenditure. This is a logical but at the same time somewhat iconoclastic point of view--iconoclastic because our habits of viewing school quality have been based upon how well schools meet input standards. This has been the methodology of school surveys and accreditation, and this has been the target of attack. It has been stated, somewhat defensively but quite accurately, that the status of the measurement art is such that it is not possible to measure all the objectives that a school is trying to attain; that the standardized test is the most sophisticated instrument but that it can measure only a part of the school effort and at the same time it has some inherent weaknesses. The position taken by National Assessment is consistent with the requirements of a good accountability tool; and if there are to be more effective and varied tools, then there must be developed in this decade better and more varied means of measuring educational output.

2. National Assessment will produce for the first time on a national basis a picture of how students and young adults are achieving in selected subject areas. The degree and distribution of knowledge and skill attainment among young people in our country should be of value in the same way that census data are in providing a picture of various facets of American

life. The early literature of National Assessment envisioned the function of assessment results in education in a way that the Gross National Product serves as an economic indicator or that certain public health measures give us a picture of the incidence of disease. We all recall that the basic motivation for the development and passage of the National Defense Education Act with its assumptions that the American schools were deficient in the teaching of mathematics and science was the fact that Russia by launching Sputnik had demonstrated fearsome superiority in space technology. This assumption was supported by articles showing how much more time was spent on Science in Russian schools than in American schools.

The point I'm making here is that there was really no dependable information on the nature and degree of achievement of young people in the United States in science.

3. A third strength of National Assessment is that it is designed to give comparisons over time. Seven subjects are to be assessed every six years and three (reading, mathematics, and science) every three years. By comparisons of results of items in one assessment of a given subject area with the same items in the next assessment of the same area, some measure of progress or lack of progress can be expressed. This will be important evidence to introduce into two kinds of arguments against the public schools: (1) children aren't learning as much or as well as they did "when I went to school", and (2) the schools are trying to do so much they aren't doing anything well.

4. A fourth strength of National Assessment has been its careful attention to the formulation of curriculum objectives which are to be evaluated. This attention is important for several reasons. First,

although it seems quite elementary, before a subject or a lesson is taught, goals should be established towards which teaching is directed; and before learning is measured, the instruments should be designed to measure the degree of success in attaining the goals. Yet the process of goal formulation itself from the classroom level to the state and national level has been a neglected process and the ordering of priorities among a list of goals is almost nonexistent. Part of the confusion is that learning objectives have been pre-determined but not always made explicitly clear by textbook writers and developers of courses of study; part is due to differences of opinion as to what is important for a child to get out of school; part is due to the rejection of certain goals by pupils and their parents; and part is the idealistic, categorical, all inclusive way that goals are expressed. In the office of the Education Commission of the States we have noted reports from various state legislative committees, interested in better accountability, calling on state departments and local school boards to study the goals they wish to meet in the schools under their jurisdiction as a first step in the evaluation-accountability process. We note, for example, this statement from the chairman of the California Legislature Joint Committee on Educational Goals and Evaluation. "It seems increasingly clear that the basic educational issue of the '70s is educational change and selection of educational goals. We sense an urgency for prompt action because most of the goal-setting processes have been rendered obsolete due to rapid, radical, and irreversible social and technological changes."

The Education Commission feels so strongly that the time has come for our schools to examine their roles and to take a look at the multitude of responsibilities thrust on them from the three R's and skill training to esthetic appreciation, to occupational training, to citizenship training,

to emotional health and the development of a positive self-concept; and to state their goals in terms of age appropriateness; and then to determine priorities among goals. The Commission also is concerned that the public and its elected representatives evaluate education as a public service among other services, such as health, welfare, highways, safety, and give it an appropriate rank among these services. Consequently, the next theme of the ECS Annual Meeting in Boston in early July will be "Educational Goals and Public Priorities."

National Assessment offers two clues in the process of improved objective formulation. The first has to do with the manner of expressing objectives and the second with the way they are developed. National Assessment from the beginning has stated that its objectives should be expressed in behavioral terms--or that they should state behaviors that students should be able to perform. You are all aware that the subject of behavioral objectives is widely debated at the present time. The debate centers around two issues: (1) whether it is in fact desirable that learning objectives be so expressed--or even, in fact, if they can be; and (2) can there be a hierarchy of behaviorally expressed objectives, or must they be so minutely expressed that each objective reflects a specific learning outcome. Advocates claim that most learning goals are so nebulous and so unrelated to specific outcomes that the goals are meaningless, and to become meaningful, they must relate to learning outcomes. Programmed instruction supporters who are among the strongest supporters of behaviorally-expressed objectives have taken the position that National Assessment objectives are still too broad in concept. National Assessment stands behind behavioral objectives because they lend themselves to more accurate measurement, and it uses its objectives as a basis for developing its measuring instruments.

A second clue from National Assessment is the way the objectives were formulated. In the subject areas being assessed it has involved three kinds of people: (1) subject matter specialists, (2) school practitioners, and (3) segments of the lay public. Panels of all three groups help formulate and review both the objectives and the exercises as to their appropriateness for an age group and as to the relatedness of objective to exercise. We think this approach is valuable in trying to bring together various points of view to bring a kind of consensus so as to eliminate confusion and antagonism from these three elements who are all interested in quality of learning in the various areas. This formula used in National Assessment has been criticized on two grounds. One is that we have not been sufficiently representative in our selection of some panels, and in some instances this may be an accurate criticism. The second is that by trying to achieve consensus on objectives and exercises, we have been bland, neutral, or non-controversial and afraid to get at cutting and controversial issues. This second issue raises a number of questions which it is not necessary to get into in this disucssion; although from philosophic and policy grounds, we in National Assessment should not ignore them.

We suggest that the formula itself has value to any school system and state department in developing for itself accepted and understood objectives and hence a degree of credibility about school programs not presently evident.

Finally, let me say that National Assessment is interested in the formulation of objectives so as to establish an evaluating methodology and not to establish a set of national goals or to develop a national curriculum.

5. A fifth strength of National Assessment is that its results are descriptive and not normative. This statement requires some explanation and some comparison with standardized tests. Out of 500 science items used with four age groups, approximately 200 are released for public scrutiny; and the percentage responses to all possible answers are published. When science is assessed again another set of items will be released and comparisons made on identical items between the two assessments. In a standardized test no items are released. A score is released and interpreted to a norm above and below which 50% of the scores fall. In National Assessment each item is a discrete measure; in standardized tests the score is based on the summing of items. In National Assessment an effort is made to use items which most of the respondents can answer correctly, some that only a small percentage can answer, and many which 40% to 60% can answer. In standardized tests items tend to group about the expectations of the large middle section of the population.

Some critics of National Assessment have charged that we are taking an undue amount of money to develop a super-test and that that is unnecessary because of the existence of large pools of items for standardized test purposes. We think that the release of items is an important facet of our project because the public can actually see the nature of the material being examined and can see what children know and can do. To stand the test of public scrutiny National Assessment applies two standards to each exercise--directionality and face validity. By directionality, it is meant that the item must bear a meaningful relationship to an objective deemed appropriate for the subject and age being examined. Exercises should not be used unless they can exemplify a stated learning objective. By content validity, it

is meant that each item stand alone on its own as a piece of information or skill that is important. Items must not be trivial. In standardized tests content validity of each item is not so important as the computations applied to the total body of items.

Some critics have charged National Assessment with faulting standardized tests. This is not so. Standardized tests perform at least two functions which assessment cannot perform. Standardized tests are valuable tools for guidance and counseling, and they can give comparative data of pupils and schools with other pupils and schools. Individual student performance compared with other students in a group, with a class or school, or with the test norm is valuable. Likewise, comparisons of schools are possible on the basis of scores. If such comparisons are sought, National Assessment cannot provide them for only student samples are used, students are not identified, and no one student takes all the items. However, with some refinements and adaptations and considerably greater expenditure, the National Assessment concept could yield certain more refined comparisons. As will be noted later, we believe that National Assessment by actually releasing items renders a valuable function. We also believe that the wider spread of item difficulty will reveal some information about learning attainments of youngsters who score very low on standardized tests. Those of us in the profession also know that the comparison of test results with established norms should be done with care depending on the characteristics of the norming population.

6. A sixth strength of National Assessment comes out of the descriptive nature of the results. While not explored in any satisfactory manner, an analysis of actual results may have some bearing on curriculum emphases

and curriculum goals. At present only results in Science, Writing, and Citizenship have been released; and only superficial reviews have been made by panels of teachers and subject matter specialists in very brief sessions. Examples of the kinds of questions raised in these panels are the following:

1) Do young people know more about the biological sciences than the physical sciences?

2) What are the implications for teaching in the apparent greater skill in social and business writing than in creative writing or the kind of writing that requires reporting on observable phenomena in one's environment?

3) Are schools doing an inferior job in teaching about the structure and function of local and state government compared to the national government?

4) Are schools really developing an understanding of what freedom of speech means in the American system?

5) Is the school really doing a job in vocational counseling?

It is the intention of the Commission to seek the reactions of educators and the public to the curriculum implications of the results in a deliberate and systematic way.

One or two reviewers in Science listed the results by surprises and disappointments--surprises that they did so well on some exercises and disappointments that they did so poorly on others. One group stated they felt the quality of writing was beyond expectation. One science reviewer suggested that physics textbooks in college might be based on higher competencies than high school graduates actually possess.

A question that I would raise relates to the accessibility of learning of all students to those kinds of knowledges and skills which are presumed to be necessary to live competently in a democratically and technically oriented society. Are general education courses of sufficient breadth and depth? Does the elective system cut many student from desired learnings? Does the concept of minimum essentials require a new look at the curriculum, graduation requirements, and the measurement of effectiveness, and, related, does our misuse of the normal curve in evaluation form a basis of rationalization for our own ineffectiveness.

7. A seventh strength of National Assessment is in its pioneering use of sampling techniques. In general a 1% random sample of ages 9 and 13 and 17 constitutes the assessees and is presumed under sampling theory to provide results, within degrees of error, comparable to those that would have been produced if the entire age group had been assessed. Likewise, out of all items prepared, samplingsof items were drawn as being representative of the objectives and difficulty levels. And finally, under the sampling approach it was possible to design the administration so that no one student would be involved more than one hour; whereas it would have taken several hours to have responded to all the exercises. A combination of the sampling technique and the use of the computer provide the technology to develop and implement a National Assessment program. Parenthetically, let me say that from the beginning National Assessment has had the services of nationally renowned statisticians in guiding this operation and also in analyzing the results by region, sex, size and type of community, social-educational level of the respondents, and by race.

The experience of National Assessment in sampling and data handling should be of value to regions, states, and large school districts who may

want to evaluate their jurisdictions for this technique should be economical in terms of cost and pupil time.

Limitations

So much for a listing of some of the strengths of National Assessment as an accountability tool. In terms of some of the characteristics of accountability as generally discussed--namely, those factors in the teaching-learning situation which have impact on quality education and what they cost--National Assessment has certain limitations.

1. First, it is readily evident that state and local results cannot be reported under the present scheme. There are both historical and financial reasons why this was not done. When National Assessment was in its early stages of conceptualization, great fear was expressed that it would lead to invidious comparisons among states and schools and that it was a subtle scheme of the federal government to take over American education and to direct it as federal planners wanted it. The design was so set that not only were there insufficiently large samples to report on a state basis but the policy position was taken that schools in the sample would not be identified publicly or their names released. It was also recognized that to have drawn sufficiently large samples for each state and to administer the exercises, score, analyze, and report on them would have required an enormous sum of money. So, consequently, National Assessment will describe learning attainment only nationally and by certain non-political sub-categories.

But taxpayers in their local communities and legislators in their states want to know about the educational quality in their jurisdictions and want accountability tools to provide hard data. After all, most of the money for schools comes from local and state resources. Therefore, it

behooves states and local systems to develop their own tools and spell out what kind of information they want. There is considerable evidence that this is being done. The Education Commission of the States in cooperation with the Education Testing Service ERIC Center will release in early April a report of state assessment development state by state.

To assist states in developing their own assessments, the ECS Steering Committee has taken two actions which will enable states to use National Assessment materials so that, if they want, state results can be compared to national results.

The first of these actions authorize the ECS to prepare and sell at cost specimen sets to state education agencies. Each set will contain booklets of released items, a manual of instructions, administrative tapes, and a scoring key. Each state education agency can duplicate the exercises, and use them as they see fit. If the state is so disposed, National Assessment materials could be incorporated into a state program or become a state program; and assuming comparable administration conditions, a state-national comparison could be made after national results have been released.

The second action would permit a state to use assessment materials at the same time a national assessment is given to its own sample, and the results can be reported simultaneously.

The Commission has ordered specimen sets for age 17 science but will delay preparation of others until the amount of interest among states is ascertained. One state is at the point of estimating costs to use National Assessment materials on a statewide sampling basis.

2. A second limitation of National Assessment is that no attempt is made to identify factors of instruction, such as class size, training of

teachers, or per pupil expenditures that may have a bearing on learning outcomes. Eventually, results will be analyzed, as indicated earlier, by region, sex, size and type of community, social-educational level, and race by black, non-black. Conceivably some of these variables could be introduced in the data gathering process and correlations made, but they are presently not in the design.

3. A third limitation is that National Assessment is restricted to 10 subject areas which are largely cognitive and skill areas. However, such limitations are characteristic of most accountability tools, notably performance contracting. National Assessment does not pressume that it is assessing all the desirable kinds of outcomes from the learning experiences that comprise the total school curriculum. It is also sensitive to the criticism that its assessment emphasis is in the cognitive domain of those subjects being assessed. Like all accountability tools, National Assessment must start with the present state of the art of educational measurement; but it intends to develop measures to get at attitudes, appreciations, and critical thinking. Some innovation in these areas is evident already, especially in the citizenship area.

This particular limitation of all accountability tools bears watching so that the tool does not determine the curriculum.

I have tried in the last few minutes to describe National Assessment as a tool of accountability and to note its strengths and its limitations. I would like to conclude by making three observations. Although National Assessment may not fulfill all the demands for an accountability tool, it has several things to contribute. First, it will for the first time give us a dependable picture of achievement in selected areas on a national basis.

I have stated this earlier, but this global view should be a baseline to note further progress. Secondly, National Assessment has within itself the capacities to add new subject areas, to develop objectives and exercises sensitive to changing demands on the schools, and the flexibility to enlarge and broaden its samples of assessees so as to enlarge its reporting categories and to add so-called input data if these seem desirable. Thirdly, it offers a model for exercise development, sampling, administration, and data analysis which could assist smaller jurisdictions in developing their own assessments.

Meanwhile, the National Assessment staff, their contractors, and their consultants feel that the problems which result in a project of this magnitude--one of the largest in social science research--and which accompany new measurement technology require a dedication to the perfection of present techniques before its design is expanded.